

1. A method for message filtering comprising the steps of:
extracting message body data from a message;
extracting message attribute data from the message;
computing a message feature vector jointly from the message body
and the message attribute data;
computing a message discriminant score using the message feature
vector; and
passing or withholding the message based on the discriminant score.
2. The method of claim 1, wherein a prompt is given to the user to
indicate when a message is passed.
3. The method of claim 2, wherein the prompt is an audio cue.
4. The method of claim 2, wherein the prompt is a visual cue.
5. The method of claim 2, wherein the prompt is an audiovisual cue.
6. The method of claim 1, wherein the message is displayed to the user
when passing.
7. The method of claim 1, wherein the message attribute features are
derived from the group comprising: message source, author, date, day of week, time
of day, corporate affiliation, and academic affiliation.
8. A method for message filtering comprising the steps of:
extracting message body data from a message;
extracting message attribute data from the message;
computing a message feature vector jointly from the message body
and the message attribute data;
computing user textual features from the user environment;
computing user attribute features from the user environment;

computing a user feature vector jointly from the user textual and attribute features;

computing a message-user similarity score from the message feature vector and the user feature vector; and

passing or withholding the message based on the message-user similarity scores.

9. The method of claim 8, wherein the message-user similarity score is computed according to the formula:

$$z_{m,u} = \frac{x_m \cdot x_u}{|x_m| \cdot |x_u|} .$$

10. The method of claim 8, wherein the user environment comprises documents currently in use and recently used documents.

11. A method for obtaining a message classifier anticipating significant events in a time series comprising:

acquiring a set of messages spanning a time period;

acquiring time series values spanning the time period;

defining significant events in the time series;

defining time intervals preceding the significant events; and

training a classifier to pass messages occurring in the defined

intervals.

12. The method of claim 11, wherein the step of training the classifier comprises:

labeling the messages occurring during the time intervals with a first

label;

labeling the remaining messages with a second label;

training the classifier to pass the first messages while withholding the

second messages.

13. The method of claim 11, wherein the significant events relate to financial time series data.
14. The method of claim 11, wherein the classifier is trained to pass messages indicative of an anticipated increase in the time series value.
15. The method of claim 11, wherein the classifier is trained to pass messages indicative of an anticipated decrease in the time series value.
16. The method of claim 11, wherein the classifier is trained to pass messages indicative of an anticipated substantially constant value for the time series value.
17. A method for filtering messages arriving in an online system, the method comprising the steps of:
 - providing a plurality of incoming messages from an online system to a user;
 - receiving an input from the user instructing the online system to act upon an incoming message;
 - labeling each incoming message in response to the instruction to act upon the incoming message to create an online labeled data set; and
 - training a classifier with the online labeled data set.
18. The method of claim 17, wherein the input from the user comprises an instruction to ignore the incoming message.
19. The method of claim 17, wherein the input from the user comprises an instruction to read and then delete the incoming message.
20. The method of claim 17, wherein the input from the user comprises an instruction to read and perform a further action upon the incoming message.

10030789-11301

21. The method of claim 20, wherein the further action comprises forwarding the message.

22. The method of claim 17, wherein the classifier is retrained at predetermined intervals with current online data sets, formed from recently received incoming messages, to provide a classifier for identifying messages of current interest to the user.

23. The method of claim 17, wherein the incoming messages are passed to the user by the classifier, and the classifier is retrained using the online labeled data set.

24. The method of claim 17, wherein the step of training a classifier comprises the steps of:

- computing feature vectors from messages;
- computing feature vectors from the user environment;
- computing a preferentially weighted message feature vector according to the formula:

$$y = [y_i] = x_m \cdot x_u \quad .$$